Name of the Scholar: Shaurya Jauhari

Name of the Supervisor: Prof. (Dr.) S.A.M. Rizvi

Department: Department of Computer Science, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi- 110025.

Title: Exegesis of Gene Expression.

Keywords: Central Dogma of Molecular Biology, Gene Expression, Functional Genomics, Microarray Data Analysis, Precision Medicine.

Abstract

There are three aspects to gene expression machinery; first, *models* have been structured to delineate the gene regulation mechanism, second are transcription and translation processes, and thirdly, the phenomenon of the central dogma of molecular biology as a whole. The key protagonists are the polymers viz. DNA, RNA, and protein.

The *technology* has been a driving force to fuel *analysis*. Foreseeing the applicability of the format of data-led-information as a result, technology is designed and adopted. The scalability and adaptation to improved underlying schema of data ensures sustenance. Microarrays have put forth an affordable way of observing instanced transcriptional spectrum. The resultant gene expression data, a snapshot of the transcriptional landscape, allows for visualizing interesting patterns of inter-operational entities. Owing to the said convenience, some arduous diseases, as an open-problem, are subjected to this technique. Cancer has been long fought to unveil suggestive causal transcripts that orchestrate abnormality in the patient. The predictions made are often generic and don't necessary comply with distinctive organismal phenotypes. Varying methods as surgical biopsy, chemotherapy, radiation therapy, even in controlled aggressive formats, have failed to contain [it] fully. Microarray datasets of a variety of cancer, that are publically available, can be studied to chase the culpable target genes. In the treatise, breast and prostate cancer stromal data is sifted to correlate with a human laterality disorder gene, *CCDC11* that is typical to affecting left-right symmetry, if mutated. The role of *WDR88* and *ARPP21* genes has been further materialized in the analysis. Another standout aspect of the study has been the associated implications of the genes in rare disorders of male breast and female prostate cancers.

Another important aspect of gene expression mechanism is the structure and flow of genomic information, stored with the genes' DNA, which is pivotal in protein synthesis.

An attempt is made in the present work to digress from the conventional understanding about the quantifiable transcripts. This novel viewing of the gene expression data hinges around deeper hierarchical insights to assign value to the sub-processes of transcription and translation. This facet of quantifying “transcription” and “translation” can divulge that the common way of using mathematical and statistical tools to determine correlation between genes in a matrix and then choosing the most potent one may not even be the way forward; rather considering some extraneous values substantively adds to the investigation. The “rate of mRNA bursts” and the “rate of protein synthesis”, as detailed ahead, are reckoned convenient.

Last but not the least is the primal aspect of all. Personalized medicine is the ultimate dispensation of all awareness about the human genome. The connectedness of all genomic information, comprehensive genes’ annotations and drugs ontology, inclusiveness of trans-demographic and epigenetic factors that orchestrate metabolism in part/full; harnessing all these aspects to provide for precise theranostics is the core of personalized medicine. Leveraging the anytime/ anywhere availability of medicinal data and access to subject/ clinician is a boon. It is therefore hypothesized, a smart card innovation- “The G-Card”, which is a manifestation of e-health and personalized electronic-health-record directives. The study explores its viability from the Indian subcontinent angle. With fast adaptations of digitization initiatives from the government for cataloguing population information, it is straightforward to allow for medical data with detailed prognostic and genetic information to represent everyone as a node to blueprinting clinical policies.

With tacit cognizance to the scalability of the problem, the current pursuit is of novel mathematical and statistical modeling techniques in cognizance to gene expression data. While deliberating the genetic mechanisms and amalgamating them with the ontological data, newfangled facets are revealed that could serve as a putative panacea to the complex gene function examinations.